MBDH:一种多尺度平衡深度哈希图像检索方法

张艺超[1]，黄樟灿[1]，陈亚雄[2,3]

(1. 武汉理工大学 理学院 数学系，武汉 430070; 2. 中国科学院西安光学精密仪器研究所，西安 710048; 3. 中国科学院大学，北京 100049)

摘　要：哈希由于其在存储和检索效率方面的优势已经被广泛用于大规模多媒体检索。通过利用数据的语义相似度来提高哈希编码质量的监督哈希近来受到更广泛关注。传统监督哈希方法将图像学习的手工特征或机器学习特征和二进制码的单独量化步骤分开，并未很好地控制量化误差，并且不能保证生成哈希码的平衡性。为了解决这个问题，提出了新的多尺度平衡深度哈希的方法。该方法采用多尺度输入，这样做有效地提升了网络对图像特征的学习效果。并且提出了新的损失函数，在很好地保留语义相似性的前提下，考虑了量化误差以及哈希码平衡性，以生成更优质的哈希码。该方法在 CIFAR-10 以及 Flickr 数据集上的最佳检索结果较当今先进方法分别提高了 5.5%和 3.1%检索精度。

关键词：多尺度；平衡性；深度哈希；卷积神经网络；图像检索

中图分类号：TP391.41　　doi: 10.3969/j.issn.1001-3695.2017.10.0962

MBDH: a multi-scale balanced deep hashing method for image retrieval

Zhang Yichao[1], Huang Zhangcan[1], Chen Yaxiong[2]

(1. *Dept of Mathematics*, *School of Science*, *Wuhan University of Technology*, *Wuhan* 430070, *China*; 2. *Xi'an Institute of Optics & Precision Mechanics*, *Chinese Academy of Sciences*, *Xi'an* 710048, *China*; 3. *University of Chinese Academy of Sciences*, *Beijing* 100049, *China*)

Abstract: Hashing has been widely used for large-scale multimedia retrieval because of its advantages of storage and retrieval efficiency. The use of the semantic similarity improving the hash coding quality has recently been more widely concerned. Traditional supervised hash methods for image retrieval represent an image as a manual feature vector or a machine learning feature vector, and then perform a separate quantization step to generate a binary code. Such methods do not control the quantization error effectively, and cannot guarantee the balance of hash code. To this end, this paper presents a new multi-scale balanced deep hash method. The method uses multi-scale input, which effectively improves the ability of learning the image features from the network. Moreover, a new loss function is proposed. Under the premise of preserving the semantic similarity, the quantization error and the balance of hash code are taken into account to generate the high quality hash code. After experimenting on two benchmark databases: CIFAR-10 and Flickr, this method has been improved by 5.5% and 3.1% of the search accuracy compared with today's advanced image retrieval methods.

Key Words: multi-scale; balance; deep hashing; convolutional neural network; image retrieval

0　引言

最近邻搜索（nearest neighbor search，NNS）[1]已经成为许多机器学习、数据挖掘、图像检索问题的基础。假设给定查询点 n，尝试找到最接近数据库中给定查询点 n 的点，这便是 NNS 的主要思想。NNS 在大数据领域的潜在应用前景受到学术界和行业的高度重视。维度灾难、存储成本和查询速度是 NNS 在处理大数据问题时所遇到的主要挑战。

哈希是 NNS 中重要而且有效的方法。可以通过构建哈希码获得非常好的效果以及实现理想的时间复杂度。哈希方法可以分为数据独立哈希及数据依赖哈希。两种方法的区别是生成哈希函数的具体方式。偏移不变核函数哈希（shift invariant kernel hashing，SIKH）[2]和最小损失哈希（minimal loss hashing，MLH）[3]是具有代表性的数据独立哈希方法，其哈希函数是人工或随机投影构建的。数据独立哈希方法的缺陷是显而易见的。理论上，过多的人工干预可能导致适应性以及准确性的缺失。

因此，从给定数据库学习哈希函数的数据依赖哈希方法被提出。这种哈希函数的构造方法可以生成更紧凑的二进制哈希

作者简介：张艺超（1992-），男（蒙古族），内蒙古乌兰察布人，硕士研究生，主要研究方向为计算机视觉、人工智能、模式识别（zhangyichao0819@163.com）；黄樟灿（1960-），男，教授，博士，主要研究方向为计算数学、图像处理；陈亚雄（1992-），男，博士研究生，主要研究方向为计算机视觉、人工智能、模式识别.

码。数据依赖的哈希方法根据给定的训练数据集是否具有标签分为监督和无监督的方法。无监督哈希函数基于一定概率理论，可以实现局部敏感效应。无监督哈希方法通过使用未标记的数据来学习哈希函数。现如今有许多无监督的哈希方法，如局部敏感哈希（locality-sensitive hashing, LSH）[2]，迭代量化（iterative quantization, ITQ）[4]和谱哈希（spectral hashing, SH）[5]。无监督的哈希算法是快速的，但是图片所含有的丰富语义没有得到很好地利用。为了避免宝贵的语义信息的丢失，监督哈希方法被提出。在某些情况下，与速度相比，精度更加被人们所重视。在这种情况下，监督哈希方法便更适用。如半监督哈希（semi-supervised hashing, SSH）[6]，最小损失哈希（minimal loss hashing, MLH）[3]，基于线性判别分析的哈希（linear discriminant analysis based hashing, LDA hash）[7]，基于内核的监督哈希（kernel based supervised hashing, KSH）[8]，基于潜在因子模型的监督哈希（latent factor models for supervised hashing, LFH）[9]。

虽然已经设计出了很多监督哈希方法，但是仍然存在以下缺陷：a）这些方法总是采用 GIST 全局特性，可能会造成语义信息丢失;b）大多数图像检索方法只能学习浅层特征，使图像信息之间的相关结构被忽略;c）一般来说，之前的方法普遍没有考虑到量化误差。

为了解决上述问题，本文提出了一种多尺度平衡深度哈希方法：

a）使用多尺度特征作为输入，这样可以得到更鲁棒的语义信息，与局部特征描述符不同，分成多尺度输入后，从深层卷积网络提取的特征向量是对整体信息进行编码的全局描述符；

b）将卷积神经网络（Convolutional Neural Network，CNN）作为深度映射引入，更多地考虑了图像信息之间的相关结构；

c）提出了一种由交叉熵、$L2$ 范数以及平衡项构成的新的损失函数。它不仅可以保留语义信息，还使得量化误差的问题得到了很好地解决，从而生成平衡且紧凑的哈希码

1　卷积神经网络哈希

卷积神经网络出现后，便快速引起了计算机视觉界的广泛关注。机器与人类在视觉感知能力方面的差距得到进一步缩小，其在物体识别，检测，图像解析和以及视频分类等各种任务中取得的显著成就推动了人工智能领域的前进。

CNN[10]是一个约束多层神经网络，其输入位于二维平面上。受人类视觉系统启发，CNN 隐藏层的神经元从上一层的局部区域获取输入，并相对于其输入区域平铺在二维特征图中。典型的 CNN 由三种类型的结构构成，即卷积层，池化层以及全连接层。卷积特征图中的神经元彼此共享权重。池化层被放置在卷积层之后，可以根据所采用的操作为最大池化还是平均池化进行分类。通过调整步幅和输入滤波器大小，卷积层和池化层都可以重叠。

CNNH[11]以及后来的改进 CNNH* [12]均为以原始图像数据为输入的两阶段框架。在第一阶段，相似度矩阵 S 被分解为一个乘积的形式:

$$S = \frac{1}{q} HH^T \tag{1}$$

其中: $\mathbf{H} \in \{-1,1\}^{n \times q}$ 表示每行是一个 q 维哈希码的近似哈希码矩阵。

在第二阶段，原始图像像素以及预生成的二进制码 **H**（CNNH*及其二进制标签 **Y**）被馈送到 CNN，CNN 的目标是使输出之间的误差最小化，并将目标二进制矢量连接到 **H** 和 **Y**。

在哈希函数学习阶段，CNNH 利用深度网络学习图像特征表示以及哈希函数。具体来说，CNNH 采用常用的深层框架作为其基本网络，并设计具有 softmax 激活的输出层以生成 q 维哈希码。CNNH 以监督的方式训练设计好的深度网络。在哈希编码学习阶段学习到哈希码被用作地面校正。另外，如果训练图像的离散类标签可用，则 CNNH 还合并这些图像标签以学习哈希函数。基于深度网络，CNNH 同时学习深层特征和哈希函数。然而，CNNH 是一个具有两个阶段的框架，其中第二阶段中的学习的深层特征不能帮助改进第一阶段中的近似哈希码学习，这极大限制了哈希学习的性能。

通常，使用具有负对数似然的逻辑回归作为损失函数用于单标签分类。也经常使用其他种类的损失函数如欧氏距离和交叉熵。本文提出了由交叉熵、$L2$ 范数以及平衡项构成的新损失函数，会在本文后面详细介绍。

最近有很多这样利用卷积神经网络学习更有效的图像表示的方法，获得了比常规哈希方法性能更好的图像表示。然而，这些方法仍存在诸多挑战。例如：图像信息没有获得更多元化的学习；在学习深层哈希算法的过程中存在不可控的量化误差，这不能最佳地兼容连续哈希码转换成离散的二进制代码，最终影响二进制代码的质量等。

2　多尺度平衡深度哈希

在相似度检索中，给出了 N 点 $\{\boldsymbol{x}_i\}_{i=1}^N$ 的训练集，每一个表示为 D 维特征向量 $x \in \mathbb{R}^D$ 。一些点与相似性标签 s_{ij} 相关联，其中 s_{ij}=1 表示 x_i 和 x_j 相似，若 s_{ij}=-1 则表示 x_i 和 x_j 不相似。目标是学习非线性哈希函数 $f: x \mapsto H \in \{-1,1\}^K$ 来对紧凑的 K 位哈希码 $H = f(x)$ 中的每个点 x 进行编码，从而保留给定对之间的相似性。

本文提出了一种新的多尺度平衡深度哈希图像检索方法。该方法通过多个尺度接受输入图像，并利用卷积神经网络哈希通道对其进行处理：a）将图片多尺度化作为输入；b）全连接哈希层，用于生成紧凑的哈希码；c）利用交叉熵函数构成的损失层来保留语义相似信息；d）利用由 $L2$ 范数构成的量化损失层来控制生成哈希码的质量。

2.1　多尺度平衡深度哈希方法

Lazebnik 等人提出了使用基于特征包（bag of feature, BoF）的方法对空间信息进行编码的空间金字塔匹配（spatial pyramid matching, SPM）方法[13]。它们代表使用几个级别或尺度的金字塔的图像。来自不同尺度的特征被组合以形成图像表示，使特

征越粗糙获得的权重越小，而特征越精细获得的权重越大。此文认为在较粗糙的层面上发现的匹配可能涉及越来越多的不同的图像特征。在本文中同样使用卷积特征图作为局部描述符，以同样思路探索多尺度情况。实验后发现卷积特征图的深层特征与传统的描述符不同：不同级别的特征的加权和不比其简单的总结表现出优越的性能。Kaiming 等人设计了一种称为空间金字塔池化（spatial pyramid pooling, SPP）[14]的方法。在 SPP 中，最后一个卷积层的特征图被划分成3或4个尺度的金字塔。首先，每个尺度的区域特征被级联，然后将比例级别特征级联到固定长度向量以被转发到下一个全连接层。在文献[15]中证明，这种策略对于无监督检索并不会取得什么好的效果，导致与其他简单组合方法相比的性能较差。

当用于图像检索时，这种特征仍然缺乏准确匹配两个图像所需的详细和本地信息。受到文献[13-15]的启发，本文更加深入地研究了应用此强大方法来获得辨别图像特征的可行性。图像由 L 级金字塔表示，并且在每个级别，图像被均匀地划分成几个重叠或非重叠区域。计算这些小区域的向量表示，然后组合区域向量以形成图像特征向量。图像的单一尺度表示仅仅是等级数 $L=1$ 的多尺度方法的特殊情况。

将小区域重新投入网络计算区域向量的时间成本将是巨大的，对于快速图像检索来说，这是不可接受的。受 Girshick 和 Tolias 等人的工作的启发[16,17]，在某一层的特征图中，原始图像区域和区域之间的线性投影。然后可以有效地计算区域特征向量，而无须重新反馈相应的图像区域。

本文提出的多尺度平衡深度哈希方法的流程图如图 1 所示。

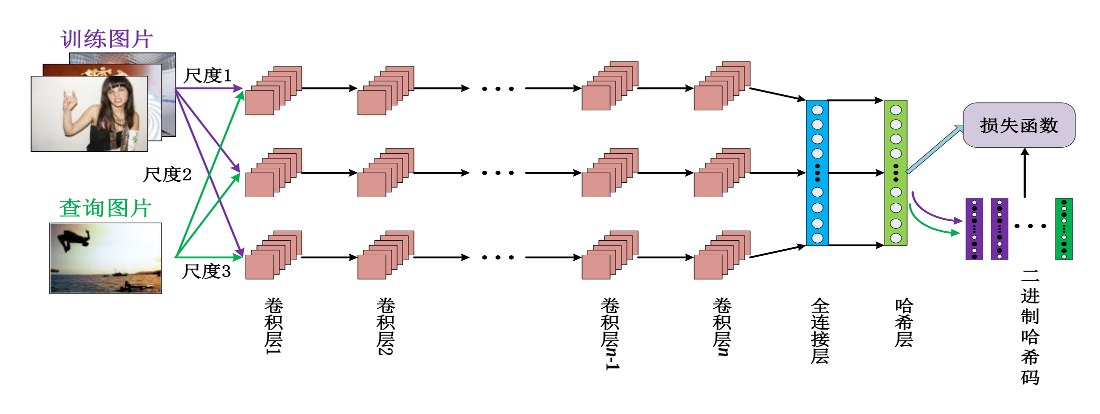


图 1　多尺度平衡深度哈希流程图

该网络在接受多尺度输入后，首先进入子卷积神经网络部分，该部分由卷积层、池化层及全连接层组成。在本文中，采用三个卷积层，第一个卷积层 32 个卷积核，大小为 3×3；第二及第三层各 64 个卷积核，大小均为 3×3。卷积层中间为两个池化层，两个最大池化的大小均为 2×2。而后连接两个全连接层，其中第二个全连接层作为哈希层。最后，进入损失函数模块，该模块由三部分组成，包括 softmax 损失部分、量化损失部分以及平衡项部分组成。

假设 $X=[x_1,x_2,x_3,...,x_N]\in\mathbb{R}^{d\times N}$ 为一个具有 N 个样本的训练集，并且 $x_n\in\mathbb{R}^d(1<n<N)$ 是第 n 个样本。哈希的最终目的是将其映射并量化为二进制编码。本文将训练样本 x_n 作为输入放入多层神经网络进行非线性变换，最终得到输出为二进制编码 $\mathbf{b}_H$。假设文中的网络是一个 L+1 层网络，分为 m 个尺度输入，其中在第 $i\in(1,2,...,m)$ 尺度下在第 $l\in(1,2,...,L)$ 层的输出可以表述为如下公式：

$$\mathbf{h}_{il}=\mathbf{W}_{il}^{T}\cdot\mathbf{h}_{i(l-1)}+\mathbf{c}_l \tag{2}$$

其中：$\mathbf{h}_{il}$ 在第 i 尺度下在第 l 层的输出，$\mathbf{W}_{il}\in\mathbb{R}^{l\times K}$ 为第 i 尺度下在第 l 层的权重，$\mathbf{c}_l$ 为该层的偏置。

各尺度在最高层的融合输出可通过如下公式求得：

$$\mathbf{H}_L=\phi\sum_{i=1}^{m}\mathbf{h}_{iL}=\phi\sum_{i=1}^{m}(\mathbf{W}_{iL}^{T}\cdot\mathbf{h}_{i(L-1)}+\mathbf{c}_L) \tag{3}$$

其中：$\mathbf{H}_L$ 为最高层的融合输出，形式为一组 K 位哈希码，ϕ 为 tanh 函数，$\mathbf{W}_L\in\mathbb{R}^{L\times K}$、$\mathbf{c}_L\in\mathbb{R}^K$ 分别为最高层的权重及偏置。

本文提出来的方法将卷积特征映射到 $[-1,1]^K$，哈希码 $\mathbf{H}_L\in[-1,1]^K$ 是连续的实值。为了获得二进制哈希码 $\mathbf{b}_H$，阈值函数为

$$\mathbf{b}_H=\mathrm{sgn}(\mathbf{H}_L) \tag{4}$$

其中：$\mathbf{b}_H\in\{-1,1\}^K$ 为一组 K 位二进制哈希码，$\mathrm{sgn}(\cdot)$ 为符号函数，如果 $x>0$，则 $\mathrm{sgn}(\cdot)=1$，否则 $\mathrm{sgn}(\cdot)=-1$。

2.2　损失函数

由于将隐层状态值转换成二进制码的离散优化是非常具有挑战性的，为了便于优化，连续松弛被应用于二进制约束，被现有的哈希方法广泛采用[18]。然而，连续松弛将产生哈希函数学习工作所广泛忽视的两个重要问题：将连续松弛代替二进制约束会带来不可控的量化误差；采用连续松弛之间的内积作为二进制代码之间的汉明距离的替代所产生的近似误差。为了控制量化误差并缩小汉明距离与其替代之间的距离，学习高质量哈希码，本文设计了一种新的损失函数，由交叉熵控制和 $L2$ 范数以及平衡项构成。它不仅可以保留语义信息，还能很好地解决量化误差的问题。

假设一张图片 $\mathcal{I}^{(n)}$ 的二进制码 $\mathbf{b}_H{}^{(n)}$ 被用作 softmax 层的输入，则预测标签 $y^{(n)}$ 的概率为

$$p(y^{(n)}=m\,|\,\mathbf{b}^{(n)})=\frac{exp(z_m)}{\sum_{j=1}^{M}exp(z_j)},m=1,2,...,M \tag{5}$$

其中：$z_m = \mathbf{w}_m^T \mathbf{b}^{(n)} + c_m$，$\mathbf{w}_m \in \mathbb{R}^{K\times 1}$ 为 softmax 层的第 m 个权重参数，c_m 为 softmax 层的第 m 个偏置参数，M 为训练图像的种类数，$\mathbf{b}^{(n)} \in \{-1,1\}^K$。

通过考虑标签 $\mathbf{Y}$ 的负对数似然，则可以获得以下优化问题：

$$\begin{aligned} \min_{\{\mathbf{B}\}} \Im_1 &= -\log p(\mathbf{Y}\,|\,\mathbf{B}) = -\sum_{n=1}^{N} \log p(y^{(n)}\,|\,\mathbf{B}) \\ &= -\sum_{n=1}^{N}\sum_{m=1}^{M} I\{y^{(n)} = m\} \log \frac{exp(z_m)}{\sum_{j=1}^{M} exp(z_j)}, \end{aligned} \tag{6}$$

其中：$I\{.\}$ 为示性函数，如果 $y^{(n)} = m$，则为 1；否则为 0。$\mathbf{B} = \{b^{(n)}\}_{n=1}^N$ 为所有图片的二进制码。上述优化问题利用标签信息去保存哈希码语义相似度。

设 $b^{(n_1)}$ 为一张图片 $\mathcal{I}^{(n_1)}$ 的二进制码，$b^{(n_2)}$ 为另一张图片 $\mathcal{I}^{(n_2)}$ 的二进制码，$b^{(n_3)}$ 是图片 $\mathcal{I}^{(n_3)}$ 的二进制码。其中图片 $\mathcal{I}^{(n_2)}$ 与图片 $\mathcal{I}^{(n_1)}$ 是相似的，图片 $\mathcal{I}^{(n_3)}$ 与图片 $\mathcal{I}^{(n_1)}$ 是不相似的。$dist_H(b^{(x)}, b^{(y)})$ 为二进制码 $b^{(x)}$ 和 $b^{(y)}$ 之间的汉明距离。优化式（8）问题可以使两张相似的图片 $\mathcal{I}^{(n_2)}$ 与 $\mathcal{I}^{(n_1)}$ 的汉明距离 $dist_H(b^{(n_1)}, b^{(n_2)})$ 尽可能小，同时使两张不相似的图片 $\mathcal{I}^{(n_1)}$ 与 $\mathcal{I}^{(n_3)}$ 的汉明距离 $dist_H(b^{(n_1)}, b^{(n_3)})$ 尽可能大[19]。

由于二元约束优化问题难以解决。在本文中，为了解决离散优化问题，提出了一种新的策略，利用连续松弛以取代二进制约束。本文采用连续哈希码代替离散二进制码。式(8) 的优化问题在可以重新定义如下：

$$\begin{aligned} \min_{\{\mathbf{H_L}, \mathbf{b_H}\}} \eta_1 &= -\sum_{n=1}^{N}\sum_{m=1}^{M} I\{y^{(n)} = m\} \log \frac{exp(\hat{z}_m)}{\sum_{j=1}^{M} exp(\hat{z}_j)} \\ s.t.\quad & \mathbf{H_L}^{(n)} = \mathbf{b_H}^{(n)}, \quad n = 1,2,3\ldots,N \\ & \mathbf{H_L}^{(n)} \in \mathbb{R}^{K\times 1}, \quad n = 1,2,3\ldots,N \\ & H_k^{(n)} \in [-1,1], \quad k = 1,2,3\ldots,K \end{aligned} \tag{7}$$

其中：η_1 为交叉熵损失函数，$\hat{z}_m = \mathbf{w}_m^T \mathbf{H}_L^{(n)} + c_m$，和 $\mathbf{H_L} = \{H^{(n)}\}_{n=1}^N$，$\mathbf{H_L}^{(n)} = \{H_1^{(n)}, H_2^{(n)}, \ldots, H_K^{(n)}\}$。

然而，连续松弛导致不可控的量化误差[20]。本文中，正则化项 $\mathcal{L}$ 被引入去控制的量化误差。使用连续哈希码和离散的二进制码之间的 $L2$ 范数作为正则化项 $\mathcal{L}$。然而，仅仅优化正则化项 $\mathcal{L}$ 就可能导致二进制码全由 1 组成。这是因为优化 $L2$ 范数项会影响哈希码的平衡性。为了维持哈希码的平衡性，利用哈希码平均值的平方作为平衡准则。这个平衡的标准鼓励哈希码每一位被映射成- 1 或 1 尽可能均匀[21]。为了产生好的二进制代码，优化问题为

$$\begin{aligned} \min_{\{\mathbf{b}_H, \mathbf{H_L}\}} \eta_2 &= \eta_1 + \gamma\mathcal{L} + \lambda\mathcal{Q} \\ &= -\sum_{n=1}^{N}\sum_{m=1}^{M} I\{y^{(n)} = m\} \log \frac{exp(\hat{z}_m)}{\sum_{j=1}^{M} exp(\hat{z}_j)} \\ &\quad + \gamma \sum_{n=1}^{N} \| \,|\mathbf{H_L}^{(n)}| - |\mathbf{b}_H^{(n)}|\, \|_2 \\ &\quad + \lambda \sum_{n=1}^{N} (\rho(\mathbf{H_L}^{(n)}))^2, \end{aligned} \tag{8}$$

其中：γ 为控制正则化强度的权重参数，λ 为控制平衡标准的相对重要性的参数，$\rho(\cdot)$ 为平均算子，$\|\cdot\|_2$ 为 $L2$ 范数，$|.|$为绝对值。正则化项 $\mathcal{L}$ 控制将连续松弛代替二进制约束产生的不可控制的量化误差。平衡的标准 $\mathcal{Q}$ 保证哈希码的每一位比特具有相同的概率为 1 或- 1，这样使得在二进制哈希码中 0 和 1 出现的几率尽可能相等。

3　实验设计及结果分析

3.1　实验算法流程

算法 1　多尺度平衡深度哈希算法

输入：训练样本 $\mathrm{I}=\{\mathcal{I}^{(n)}\}_{n=1}^N$ 和它们对应的标签向量 $Y=\{y^{(n)}\}^N$。

输出：所有权重参数 W；所有偏置参数 b。

初始化：权重采用高斯分布初始化

循环：

1：通过前向传播计算 h_{il}；

2：根据式(2)(3)计算哈希码 $H_L(x_i)$；

3：根据式(4)计算预测输出 $\hat{z}_i$；

4：根据式(8)利用 $W, H_L(x_i), z_i, \hat{z}_i$ 来计算 η_2；

5：利用来随机梯度下降法（SGD）更新参数 W, b；

直到：达到固定迭代次数

返回：W, b

3.2　实验数据集

根据本文所提出的方法，利用两个基准数据集 CIFAR-10，Flickr 对本方法进行评估。

a)CIFAR-10 图像数据库。这个数据集是 8 000 万张 Tiny 图像数据集的子集，其中分为 10 类 (每类 6 000 张图像) 的对象的彩色图像。每个图像的大小是 32×32。这些类包含飞机、汽车、鸟、猫、鹿、狗、青蛙、马、船和卡车。对于 CIFAR-10 数据集，从每个类别中随机选择 1000 张图像以形成测试查询集，并且剩余的 50 000 张图像用作训练集。根本文使用 512 维 GIST 特征作为传统特征表示和 4 096 维 CNN 特征作为深度语义特征表示。

b)Flickr 图像数据库。由 Flickr 收集的 25000 张图像组成，其中每个图像都标有一个语义概念，共 38 种语义概念。在本文中，将该子集的图像调整为 32×32。

3.3　实验环境配置

本文实验环境为：GeForce GTX Titan X GPU、中央处理器为 Intel(R) Core i7-5930K 3.50 GHz、内存为 64 GB、操作系统为 Ubuntu 14.04。所提出的模型使用开源库 KERAS 来实现。模型总体目标函数通过随机梯度下降 (SGD) 优化。SGD 的学习率为 10-6，每次更新后的学习率衰减值为 10-7，Nesterov 动量为 0.9。每批次的大小为 32。将本文网络的输入层和隐藏层之间的初始权重设置为正态分布。其初始循环权重矩阵设置为单位矩阵，剩余权重采用高斯分布。在本实验中，参数 γ 和 λ 分别设定为 0.001 和 0.001。本文后面将会进行针对这些参数配置的实验，说明参数选择的原因及其合理性。

3.4　实验评价指标

为了评估哈希方法的有效性，考虑了普遍用于定量性能比较的几个指标后，最终采用如下三个度量:

a）平均检索精度 (MAP, Mean Average Precision),它是每个查询样本平均准确率的平均值。

b）以哈希码长度 48bits 作为参考时，精度@前 *n* 个返回结果 (Precision curve w.r.t. top-n @ 48 bits)。根据查询样本和数据集之间的汉明距离对数据库中的图像进行排名，并计算排名列表的前 *n* 个精度。

c）汉明半径 2 内的精度，计算查询样本和数据集之间的汉明距离小于 2 的精度。

在实验中，本文应用类标签作为正确的标准。通过检查查询图像和返回的图像是否具有相同的标签来计算所有指标。如果这些指标的值更高，性能更好。

3.5 实验结果及分析

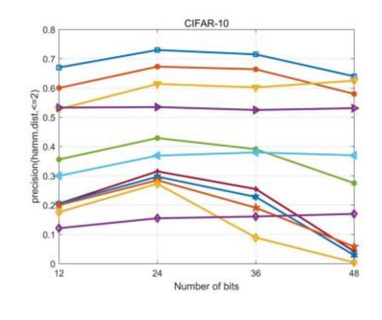
1）CIFAR-10 数据库上的结果

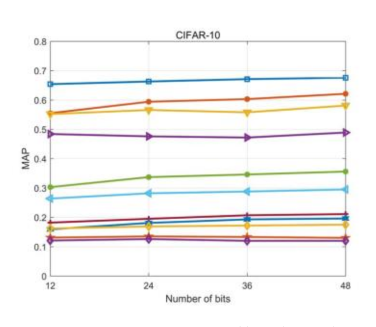
本文利用上述三个指标评估了提出的多尺度卷积神经网络哈希方法的检索质量，并与经典方法以及一些当今前沿方法进行了比较，其中包括三个无监督方法：LSH[2]、SH[5]和 ITQ[4]方法；以及八个监督方法：DHN[22]、CNNH[11]及其变体 CNNH*[12]、DNNH[12]、KSH[8]、MLH[3]、BRE[23]和 ITQ-CCA[24]方法。LSH、SH、ITQ、KSH、MLH、BRE 和 ITQ-CCA 是传统的哈希方法，它们使用 512 维向量作为输入来学习哈希函数。其他四种哈希方法 (即 DNH、DNNH、CNNH*和 CNNH) 使用 4096 维 CNN 特征作为输入来执行哈希函数。

可以清楚地发现：a)比较利用手工特征和利用 4096 维 CNN 特征的相同的方法，可以看出 CNN 特征可以提高传统方法的检索精度;b)与其他深度哈希方法 CNNH、CNNH*、DNNH、DHN 相比，提出的 MBDH 将检索平均 MAP 从 42.9% (CNNH)、48.4% (CNNH*)、55.2% (DNNH)、55.5%(DHN)至 67.6%。这是因为所提出的方法运用了多尺度的图片输入，并且设计了新的损失函数，其保持了哈希码的语义相似性和平衡性，并且同时考虑了连续编码转换为离散二进制码产生的量化误差，从而可以提高检索精度。

表 1 CIFAR-10 数据库上 MAP 的图像检索结果

方法	CIFAR-10 (MAP)			
	12bits	24 bits	32 bits	48 bits
MBDH	**0.654**	**0.663**	**0.671**	**0.676**
DNH	0.555	0.594	0.603	0.621
DNNH	0.552	0.566	0.558	0.581
CNNH*	0.484	0.476	0.472	0.489
CNNH	0.429	0.511	0.509	0.522
KSH	0.303	0.337	0.346	0.356
ITQ-CCA	0.264	0.282	0.288	0.295
MLH	0.182	0.195	0.207	0.211
BRE	0.159	0.181	0.193	0.196
SH	0.131	0.135	0.133	0.130
ITQ	0.162	0.169	0.172	0.175
LSH	0.121	0.126	0.120	0.120





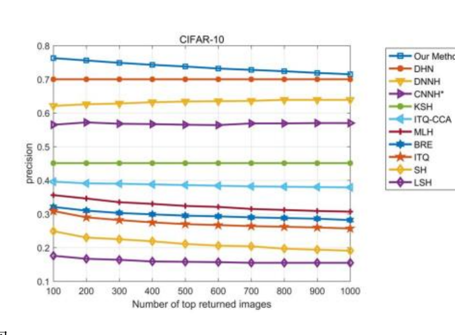


图 2 CIFAR-10 数据集的结果图

图 2 左边显示了在 CIFAR-10 数据集上的不同比特情况下汉明距离小于等于 2 的精度曲线，可以看出，所提出的 MBDH 在所有比特的汉明距离小于等于 2 实现了最佳的检索精度。图 2 中间展示了所提出的 MBDH 与其他方法在所有比特情况下前的 MAP。图 2 右边显示了在 CIFAR-10 数据集上 48 bit 情况下前 1000 张返回图片的精度曲线，所提出的 MBDH 仍然实现了最佳的检索精度。

MBDH 在本数据集上查询 10 个类返回的前 10 个结果如图 3 所示。其中，左侧为查询图片，右侧有红色方框标记的为错误结果。



图 3 MBDH 在 CIFAR-10 数据集查询返回前 10 个结果

2）Flickr 数据库上的结果

在该实验中采用的实验设置与上文相同。表 2 表示在 Flickr 数据集上不同长度的哈希码的检索 MAP 结果。

表 2　Flickr 数据库上 MAP 的图像检索结果

方法	Flickr (MAP)			
	12bits	24 bits	32 bits	48 bits
MBDH	**0.853**	**0.859**	**0.863**	**0.872**
DHN	0.810	0.828	0.829	0.841
DNNH	0.783	0.789	0.791	0.802
CNNH*	0.749	0.761	0.768	0.776
CNNH	0.732	0.734	0.741	0.740
KSH	0.690	0.702	0.702	0.706
ITQ-CCA	0.513	0.531	0.540	0.555
MLH	0.610	0.618	0.629	0.634
BRE	0.571	0.592	0.599	0.604
SH	0.531	0.533	0.531	0.529
ITQ	0.544	0.555	0.560	0.570
LSH	0.499	0.513	0.521	0.548

可以清楚地看到：

a）利用 4 096 维 CNN 特征的传统哈希方法比使用手工特征的相同方法具有更好的性能，可以发现 CNN 特征可以提高传统方法的检索精度。

b）所提出的 MBDH 实现了优于其他比较的现有哈希技术方法的性能。并且所提出的 MBDH 将平均 MAP 从 DHN 的 84.1% 改进到 87.2% 。

c）所提出的 MBDH 可以实现优于其他 8 种监督哈希方法 (即 DNH、DNNH、CNNH、CNNH*、KSH、MLH、BRE 和 ITQ-CCA 方法) 的性能，因为其使用多尺度输入，并使用了新的损失函数。

3.6　关于参数配置影响的实验分析

将两个参数 γ 和 λ 设置为从 10^{-5} 到 1 的不同值，为了方便记录，此处 γ 和 λ 取相同的值。并计算长度为 48 位的哈希码在 CIFAR-10 数据集上的 MAP。结果如表 3 所示。

表 3　参数配置影响实验结果

评价度量	不同参数 γ 和 λ 取值下的实验结果(48bits)					
	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}	1
MAP	0.632	0.658	0.676	0.665	0.643	0.60

从表 3 可以明显观察到当参数 γ 和 λ 取 10^{-3} 时，MAP 达到最高值为 0.676。解释了本文参数选择的原因及合理性。

3.7　多尺度有效性的实验分析

为了进一步验证本文所提出的多尺度平衡深度哈希方法的有效性，本文还对相同网络结构的单尺度以及双尺度方法进行了实验，计算三种方法在哈希码长度为 48 位的情况下三个评价指标的值。结果如表 4 所示。

从表 4 中可以明显看出，本文提出的方法的三个度量指标值都明显高于单尺度及双尺度方法。尤其较单尺度方法，MAP 提高了 4.4%。通过上述实验，可以进一步证明本文提出的 MBDH 的有效性。

表 4　多尺度输入有效性实验结果

方法	评价度量		
	MAP	Precision @500	Hamm dist<=2
单尺度	0.632	0.6092	0.6174
双尺度	0.661	0.6423	0.6521
多尺度	0.676	0.6538	0.6624

4　结束语

本文提出一个简单而有效的多尺度平衡深度哈希模型 MBDH 来学习二进制哈希码，用于快速图像检索。该方法不依赖于数据的对称相似性，提出的深度哈希网络架构将单一尺度输入换为多尺度输入，并同时优化了语义相似性的交叉熵损失以及生成紧凑哈希码时产生的量化损失，并考虑到了哈希码的平衡性。本文进行了广泛的实验，并提供了 MBDH 在两个基准数据库上的实验结果以及与多种当今先进哈希方法的比较评估。实验结果表明，通过采用多尺度输入，并对损失函数进行优化后，MBDH 分别对 CIFAR-10 以及 Flickr 数据集的最佳检索结果较当今先进方法分别提高了 5.5%和 3.1%检索精度。进一步展示了所提出的方法对数据量在 100 万以上的大规模数据集的可扩展性和有效性。

参考文献：

[1] 易唐唐，黄立宏. CBIR 中一种基于最近邻的改进相关反馈算法 [J//OL]. 计算机应用研究, 2015, 32 (08): 2326-2330.

[2] Raginsky M, Lazebnik S. Locality-sensitive binary codes from shift-invariant kernels [C]// Advances in Neural Information Processing Systems. 2009: 1509-1517.

[3] Norouzi M, Blei D M. Minimal loss hashing for compact binary codes [C]// Proc of International Conference on Machine Learning. 2011: 353-360.

[4] Gong Y, Lazebnik S, Gordo A, et al. Iterative quantization: a procrustean approach to learning binary codes for large-scale image retrieval [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2013, 35 (12): 2916-2929.

[5] Weiss Y, Torralba A, Fergus R. Spectral hashing [C]// Advances in Neural Information Processing Systems. 2009: 1753-1760.

[6] Wang J, Kumar S, Chang S F. Semi-supervised hashing for scalable image retrieval [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2010: 3424-3431.

[7] Strecha C, Bronstein A, Bronstein M, et al. LDAHash: improved matching with smaller descriptors [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2012, 34 (1): 66-78.

[8] Liu W, Wang J, Ji R, et al. Supervised hashing with kernels [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2012: 2074-2081.

[9] Zhang P, Zhang W, Li W J, et al. Supervised hashing with latent factor models [C]// Proc of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval. 2014: 173-182.

[10] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [C]// Advances in Neural Information Processing Systems. 2012: 1097-1105.

[11] Xia R, Pan Y, Lai H, et al. Supervised hashing for image retrieval via image representation learning [C]// Proc of the 28th AAAI Conference on Artificial Intelligence. 2014: 2156-2162.

[12] Lai H, Pan Y, Liu Y, et al. Simultaneous feature learning and hash coding with deep neural networks [C]// Proc of International Conference on Computer Vision and Pattern Recognition. 2015: 3270-3278.

[13] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories [C]// Proc of International Conference on Computer Vision and Pattern Recognition. 2006: 2169-2178.

[14] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2015, 37 (9): 1904-1916.

[15] Hao J, Dong J, Wang W, et al. What Is the Best Practice for CNNs Applied to Visual Instance Retrieval? [J]. arXiv preprint arXiv: 1611. 01640, 2016.

[16] Girshick R. Fast r-CNN [C]// Proc of International Conference on Computer Vision and Pattern Recognition. 2015: 1440-1448.

[17] Tolias G, Sicre R, Jégou H. Particular object retrieval with integral max-pooling of CNN activations [J]. arXiv preprint arXiv: 1511. 05879, 2015.

[18] Wang J, Shen H T, Song J, et al. Hashing for similarity search: A survey [J]. arXiv preprint arXiv: 1408. 2927, 2014.

[19] Li W J, Wang S, Kang W C. Feature learning based deep supervised hashing with pairwise labels [J]. arXiv preprint arXiv: 1511. 03855, 2015.

[20] Kang W C, Li W J, Zhou Z H. Column Sampling Based Discrete Supervised Hashing [C]// Proc of the 30th AAAI Conference on Artificial Intelligence. 2016: 1230-1236.

[21] Do T T, Doan A Z, Cheung N M. Discrete hashing with deep neural network [J]. arXiv preprint arXiv: 1508. 07148, 2015.

[22] Zhu H, Long M, Wang J, et al. Deep Hashing Network for Efficient Similarity Retrieval [C]// Proc of the 30th AAAI Conference on Artificial Intelligence. 2016: 2415-2421.

[23] Kulis B, Darrell T. Learning to hash with binary reconstructive embeddings [C]// Advances in Neural Information Processing Systems. 2009: 1042-1050.

[24] Gong Y, Lazebnik S, Gordo A, et al. Iterative quantization: a procrustean approach to learning binary codes for large-scale image retrieval [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2013, 35 (12): 2916-292